3-4.

a.

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = 3086/11 = \$280{,}545.50$$

To compute the median, rank the observations and find the middle value.

112 264 271 287 289 293 298 305 317 325 325

Median = $293,000
Mode = $325,000
Student answers will vary. The distribution somewhat left skewed due to the low value of 112. Either median or mean could be selected.

b. the 1st quartile is equal to the 25th percentile

$$i = \frac{p}{100}(n+1) = (25/100)(12) = 3 \text{ or } 3^{rd} \text{ observation} = \$271{,}000$$

the 3rd quartile is equal to the 75th percentile

$$i = \frac{p}{100}(n+1) = (75/100)(12) = 9 \text{ or 9th observation} = \$317{,}000$$

3-14.

a.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 456/24 = 19$$

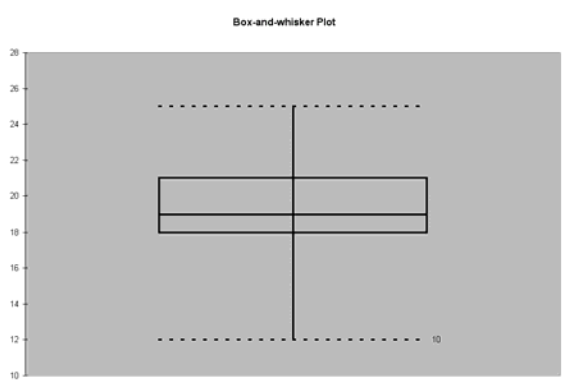
To compute the median, rank the observations and find the average of the middle two values.

10 12 14 14 17 17 18 18 19 19 19 19
19 20 20 21 21 21 21 22 22 23 25 25

Median = (19 + 19)/2 = 19
Mode = 19

b. This data is symmetrical since the mean = median = mode c.

Box-and-whisker Plot	
Five-number Summary	
Minimum	12
First Quartile	18
Median	19
Third Quartile	21
Maximum	25



The box plot does support the idea that the distributions are symmetric although the median is not directly in the center between the Q1 and Q3.

3-18

a. Sorting the data and determine what position 45,000 is in you can solve the percentile equation for the percentile.

$i = \frac{p}{100}(n+1)$ = (75/100)*(200+1) = 150.75 ; This 75th percentile can be approximated as a value somewhere between the 150th and 151st value in the data. You can use PHStat's Stack feature under Data

148	44,879
149	44,879
150	44,904
151	44,980
152	45,052
153	45,148
154	45,153
155	45,227
156	45,228
157	45,276

Preparation to reorganize the data. Then sorting the data you get the following:

Thus, the 75th percentile is a value between 44,904 and 44,980. It would be .75 of the distance between these two values up from 44,904 which gives 44,961. Note, Excel gives a slightly different value of 44,923 since it uses a slightly different methodology for interpolating the difference between the 150th and 151st data values.

b. Using Excel's Average and Median functions you can find that
Mean = 42,261
Median = 42,326

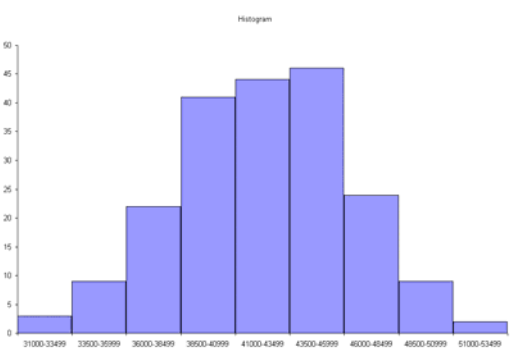
c. Using Excel's histogram tool the following histogram can be created. According to Sturges' rule, the appropriate number of classes should be:

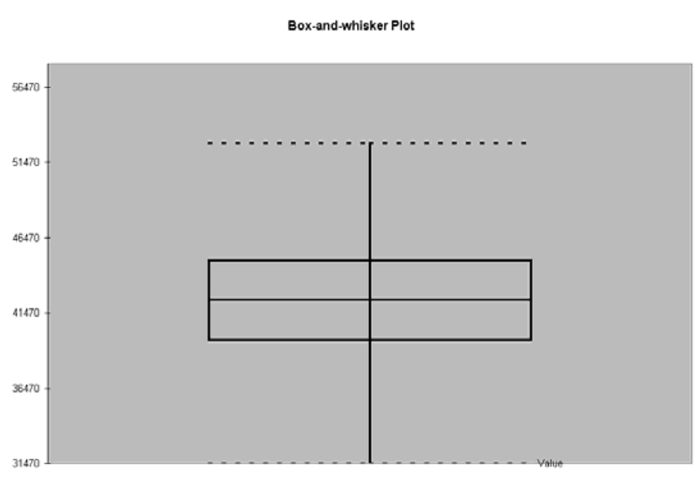
$$1 + 3.322(\log(200)) = 8.6$$ which is rounded to 9 classes.

The class width is:

$$w = \frac{High - Low}{\#} = \frac{52{,}774 - 31{,}476}{9} = \frac{21{,}298}{9} = 2{,}366.44$$

We choose to round this to 2,500 and start at 31,000 giving the following histogram:





Histograms and box and whiskers plots have certain things in common. In both instances, we get an idea of how the data are distributed, where the center is, and what the shape of the distribution is and how spread out the data are. The histogram breaks the data down into classes and illustrates the actual number of values in each class where the box and whiskers plot shows the median and the inter-quartile range.

d. Student answers will vary.

3-20.

The following parameters are computed assuming the data represent a population:

Range = High – Low = 24 – 9 = 15

Variance = $\sigma^2 = \frac{\sum_{i=1}^{N}(x-\mu)^2}{N} =$

$$\frac{(16-16.4375)^2+(23-16.475)^2+...(14-16.4375)^2}{16} = 19.62$$

Standard Deviation = $\sigma = \sqrt{\sigma} = \sqrt{19.62} = 4.43$

3.36

X	X - $\bar{x}$	$(X - \bar{x})^2$
10	-6.1	37.21
19	2.9	8.41
17	0.9	0.81
19	2.9	8.41
12	-4.1	16.81
20	3.9	15.21
20	3.9	15.21
15	-1.1	1.21
16	-0.1	0.01
13	-3.1	9.61
161		112.9

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ = 161/10 = 16.1

$S^2 = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}$ = 112.9/(10-1) = 12.5444

$S = \sqrt{S^2} = \sqrt{12.5444}$ = 3.5418

16.1 + (1)3.5418 = 19.6418 or > 19 visits are required

3-33.

Excel's descriptive statistics tool can be used to determine the answers to this problem. You can use the descriptive statistics tool to determine the interquartile range by inputting the positions but the answer may be slightly different than doing it by hand using the equations shown in the text.

a.

Credit Card Account Balance	
Mean	753.68
Standard Error	17.00202273
Median	737
Mode	600

Standard Deviation	294.4836719
Sample Variance	86720.63304
Kurtosis	-0.517456441
Skewness	0.113822027
Range	1394
Minimum	99
Maximum	1493
Sum	226104
Count	300
Largest(75)	974
Smallest(75)	544

The interquartile range is the Largest (75) from the table above minus the Smallest (75) from the table above. 974 – 544 = 430 Note the largest 75 corresponds to Q3 (3^{rd} Quartile) and the smallest 75 corresponds to Q1 (1^{st} quartile).

b. For Males:

Credit Card Balances - Male	
Mean	746.512931
Standard Error	19.33632279
Median	738.5
Mode	1018
Standard Deviation	294.5220941
Sample Variance	86743.2639
Kurtosis	-0.605714694
Skewness	0.085179909
Range	1344
Minimum	99
Maximum	1443
Sum	173191
Count	232
Largest(58)	960
Smallest(58)	538

The interquartile range is the Largest (58) from the table above minus the Smallest (58) from the table above. 960 – 538 = 422

For Females:

Credit Card Account Balance - Female	
Mean	778.1323529
Standard Error	35.80014705
Median	737
Mode	600
Standard Deviation	295.2155754
Sample Variance	87152.23595
Kurtosis	-0.199115911
Skewness	0.219080607
Range	1358
Minimum	135
Maximum	1493
Sum	52913
Count	68
Largest(17)	990
Smallest(17)	587

The interquartile range is the Largest (17) from the table above minus the Smallest (17) from the table above. 990 – 587 = 403

c. Student answers will vary.

3-34.

a.

X	$X - \bar{x}$	$(X - \bar{x})^2$
16	-0.69231	0.47929
23	6.307692	39.78698
17	0.307692	0.094675
24	7.307692	53.40237
9	-7.69231	59.1716
11	-5.69231	32.40237
13	-3.69231	13.63314
15	-1.69231	2.863905

15	-1.69231	2.863905
23	6.307692	39.78698
18	1.307692	1.710059
16	-0.69231	0.47929
17	0.307692	0.094675
217		246.7692

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 217/13 = 16.692$$

$$S^2 = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n-1} = 246.7692/(13\text{-}1) = 20.564$$

$$S = \sqrt{S^2} = \sqrt{20.564} = 4.535$$

b.

$$CV = \frac{S}{\bar{x}}(100) = (4.535/16.692)(100) = 27.17\%$$

Coefficient of variation is used to measure the relative variation in the data. It is used most to compare two or more data sets when the means of the two are different

c. Using Tchebysheff's Theorem at least 75% of the data should fall within 2 standard deviations of the mean. For this data that would be

$16.692 \pm 2(4.535)$

7.622 – 25.762

100% of the data fell within this range so you could say the range is in fact conservative.

d. Using the empirical rule approximately 68% of the data should fall within 1 standard deviation of the mean. For this data that would be

$16.692 \pm (1)(4.535)$

12.157 – 21.227

8 values out of 13 fell within that range or 61.5%

3-46.

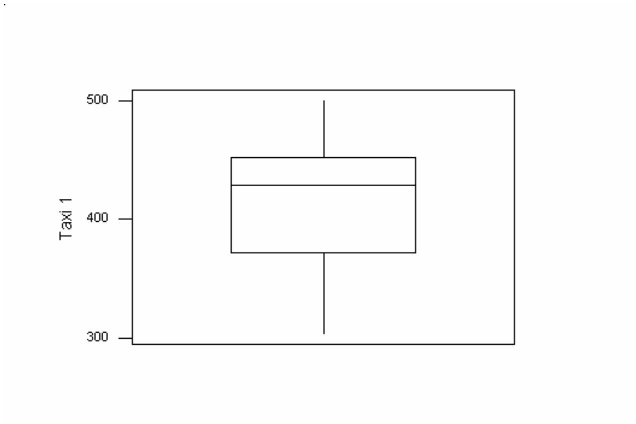
a. Student answers will vary but below is the frequency distribution if you decide to use 10 classes, for example. The Frequency function in Excel can be used.

Classes	*Frequency*
271 - 300	1
301 - 330	4
331 - 360	18
361 - 390	31
391 - 420	36
421 - 450	34
451 - 480	19
481 - 510	10
511 - 540	6
541 - 570	1

b. Below is the frequency distribution from the standardized scores using a mean of 415.3 and a standard deviation of 49.808

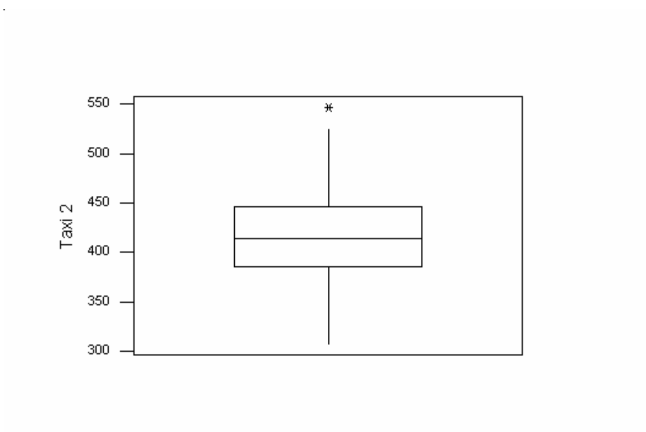
Classes	*Frequency*
-1.95	3
-1.4	10
-0.85	23
-0.3	23
0.25	39
0.8	33
1.35	14
1.9	8
2.45	5
3	2

c. Use either Minitab or PHStat to create the Box and Whiskers plots

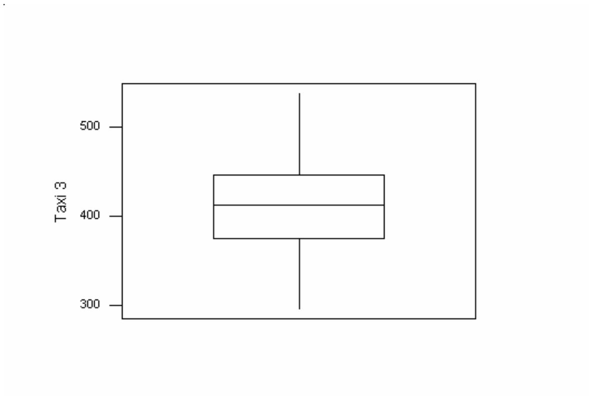


Taxi 1

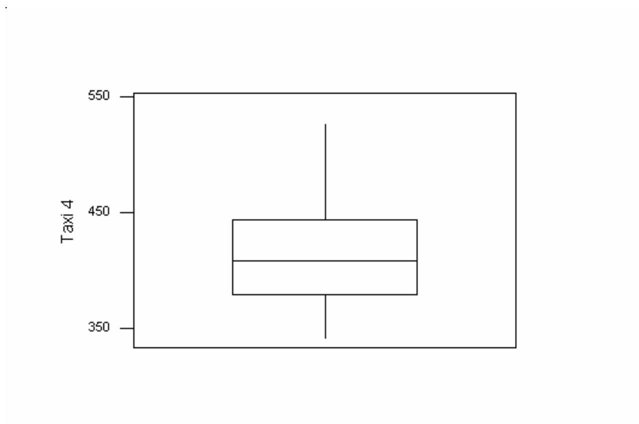
Taxi 2:



Taxi 3:



Taxi 4:



Only taxi 2 shows an outlier observation.

3-48.

Use Excel's average and standard deviation functions to determine the mean and standard deviation of each type of bread. The results are shown below.

a. White Bread has the highest average daily demand.

Bread Type	Average
White	599.77273
Wheat	530.40909
Multigrain	470.36364
Black	383.59091
Cinnamon Raisin	139.72727
Sour Dough French	127.09091
Light Oat	261.63636

b. Use Excel's Histogram feature to develop the frequency distribution for each bread type. Student answers will vary depending on number of classes selected and class widths used, but shown below are the results if you let Excel set up the bins.

White Bread	*Frequency*
251 - 375	1
376 - 500	5
501 - 625	7
626 - 750	5
751 - 875	4

Cinnamon	*Frequency*
54.76 - 84.00	1
84.01 - 113.25	4
113.26 - 142.50	8
142.51 - 171.75	4
171.76 - 201.00	5

Wheat Bread	*Frequency*
264.26 -	1
352.01 -	3
439.76 -	8
527.51 -	4
615.26 -	6

Sour Dough	*Frequency*
64 - 88	1
89 - 113	8
114 - 138	4
139 - 163	7
164 - 188	2

Multigrain	*Frequency*
212.76 -	1
299.01 -	4
385.26 -	8
471.51 -	2
557.76 -	7

Light Oat Bread	*Frequency*
127 - 172	2
173 - 218	3
219 - 264	8
265 - 310	4
311 - 356	5

Black Bread	*Frequency*
182.76 -	1
256.01 -	5
329.26 -	7
402.51 -	6
475.76 -	3

c. White Bread has the highest standard deviation.

Bread Type	**Standard**
White	149.0550975

Wheat	107.0236552
Multigrain	108.0130263
Black	81.9510069
Cinnamon Raisin	33.01973698
Sour Dough French	28.68367561
Light Oat	57.89526295

d. Use Excel to calculate the coefficient of variation for each bread type. White bread has the greatest relative variability and wheat bread has the lowest relative variability.

Bread Type	Coefficient of Variation
White	24.85%
Wheat	20.18%
Multigrain	22.96%
Black	21.36%
Cinnamon Raisin	23.63%
Sour Dough French	22.57%
Light Oat	22.13%

e. Use Tchebysheff's Theorem to calculate the upper range of two standard deviations from the mean. You must use Tchebysheff's Theorem because you do not know if the data is bell-shaped.

Bread Type	Required Loaves
White	897.8829
Wheat	744.4564
Multigrain	686.3897
Black	547.4929
Cinnamon Raisin	205.7667
Sour Dough French	184.4583
Light Oat	377.4269

f. Use Excel's Pivot Table feature to calculate the average total loaves by day of week. The highest average is on day 6.

Average of Total Loaves Sold	

Day of Week	Total
1	2196
2	2947
3	2388
4	2335.5
5	2336.8
6	3116.5
8	1772
Grand Total	2512.590909

3-54.

The standard deviation measures how the data is spread around the mean. If the means are not the same then comparing the measure of the spread around the mean does not give any useful comparison of two data sets. The appropriate measure is the coefficient of variation. This essentially standardizes the data so that a comparison between two data sets is meaningful. The larger the coefficient of variation the more variable is the data. You would then be able to say that one data set has a larger relative variation than another data set.

3-59.

Note, even though the company surveyed 30 families, only 15 surveys were returned.

X	X - $\bar{x}$	(X - $\bar{x}$)^2
38	13.26667	176.004444
44	19.26667	371.204444
11	-13.7333	188.604444
26	1.266667	1.60444444
19	-5.73333	32.8711111
13	-11.7333	137.671111
45	20.26667	410.737778
27	2.266667	5.13777778
11	-13.7333	188.604444
19	-5.73333	32.8711111
19	-5.73333	32.8711111
26	1.266667	1.60444444
20	-4.73333	22.4044444
19	-5.73333	32.8711111
34	9.266667	85.8711111

371 1,720.9333

a. The mean measures the numerical center of the data by summing the values and dividing by the number of observations.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 371/15 = 24.7333$$

b. To compute the median, rank the observations and select the middle value since the number of observations is odd.

11 11 13 19 19 19 19 20 26 26 27 34 38 44 45

Median = 20

The median < mean which means the data is skewed right.

c. The mode is 19

d. The standard deviation is essentially an average of how the data is spread around the mean.

$$S^2 = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n-1} = 1720.9333/(15\text{-}1) = 122.9238$$

$$S = \sqrt{S^2} = \sqrt{122.9238} = 11.0871$$

e. The 1st quartile is equal to the 25th percentile

$$i = \frac{p}{100}(n+1) = (25/100)(15+1) = 4 \text{ or } 4^{th} \text{ observation} = 19$$

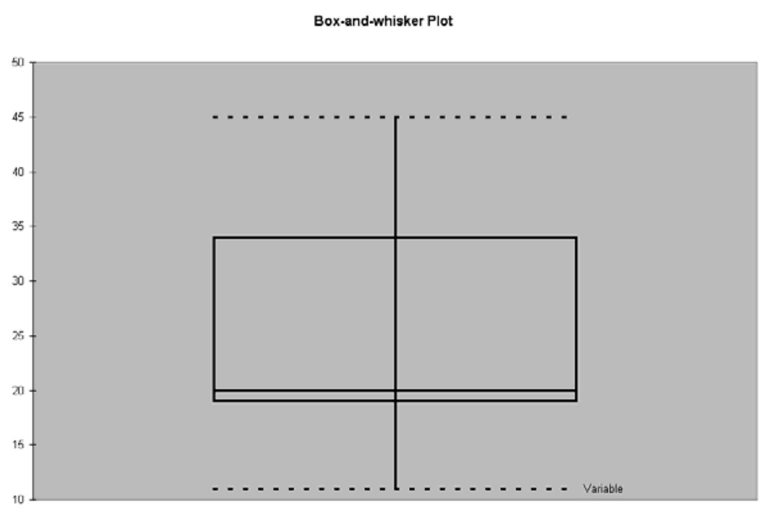
The 3rd quartile is equal to the 75th percentile

$$i = \frac{p}{100}(n+1) = (75/100)(15+1) = 12 \text{or } 12^{th} \text{ observation} = 34$$

Interquartile Range = 34 - 19 = 15

The interquartile range is often preferred to the range because if you have outliers it will likely minimize the effect of the outliers because you do not use the extreme outside values since you are only looking at the middle 50% of the values.

f.



The box plot shows that the distribution is not symmetrical since the median line is not centered in the box but instead is located very close to the Q1 value.